MATHEMATICA
Policy Research, Inc.

600 Maryland Ave., SW, Suite 550
Washington, DC 20024-2512
Telephone (202) 484-9220
Fax (202) 863-1763
www.mathematica-mpr.com

MEMORANDUM

TO: Anne Rothbaum

FROM: Duncan Chaplin, Minki Chatterji, and Denzel Hankinson **DATE:** 7/2/2009

SUBJECT: Final Design Report for the Millennium Challenge Corporation (MCC) Impact Evaluation Design & Implementation Services—Tanzania Energy Sector Project

---

A. INTRODUCTION

This design report for an evaluation of three MCC-funded electricity projects is an update of the report MPR submitted to MCC on May 6 2009. This update incorporates two new features:

- A flexible design option that would enable us to finalize the design after conducting a baseline survey
- A description of how we might go about choosing which method will be used to estimate the impacts of line extensions for the D&E project

In addition we have updated the following sections.

- Power analysis to align with the new design option and the sample sizes given in the Terms of Reference (TOR) that we submitted to MCC on June 16, 2009
- Critical assumptions
- Next steps
- Study timeline

These changes are mostly in Appendix B of this document. While this document is the Final Design Report for Phase I of this project, we note that additional information and design work will be needed in order to conduct an evaluation of the MCC-funded electricity projects in Tanzania. In particular we need to know which subprojects will be funded and when they will be implemented. Our study could also benefit greatly from having census data on the numbers and characteristics of households by Primary Sampling Unit (PSU) in the areas that will be covered in our study. All of this information could be used to update the power calculations and finalize the study design at a later time.

MCC is funding the following electrification projects in Tanzania:

1. The Cable project, which involves the installation of a new electricity cable from the mainland to Unguja Island in Zanzibar
2. The Hydropower project, which will extend the distribution network to additional towns and villages in the Kigoma Region
3. The Distribution Systems Rehabilitation and Extension Activity (D&E) project, which involves the rehabilitation and extension of existing electrical lines in six regions in Tanzania

The two main questions for our evaluation are:

1. Does access to electricity lead to increased household income and better health and education outcomes and to what extent?
2. Does access to electricity lead to increased business activity, including new firms, capital investments, and greater levels of employment and to what extent?

By access we mean getting an electric line extension funded by MCC, having a subsidy to connect to one of these new lines, and/or actually being connected. Our research designs are well designed for estimating impacts of the first two types of access. Estimating the impacts of being connected may be somewhat more difficult as it may be hard to separate the effects of being connected from other changes that are associated with extending lines and providing subsidies. In particular, extending lines may cause local governments to improve roads and water systems. Similarly, subsidies may provide direct monetary benefits to some households, in addition to affecting electricity use.

In addressing the key questions above, the evaluation will, when possible, also address (1) differences in impact of the program, by gender, age, and income and (2) unintended results of the program. A conceptual framework illustrating our approach is in Figure 1. In addition, Table 1 outlines the projects, our evaluation questions, and possible study approaches.

We will conduct separate studies for the Cable and D&E projects.[1] Our plans for the Cable project are preliminary. For that project, we may use an interrupted time series (ITS) method and/or a pre-post cost (PPC) method, both focused on business-related outcomes. For the D&E study we may use regression discontinuity (RD). However, as explained below, we have concerns about the feasibility of this option, so we are considering other options as well. Alternatives include using the difference-in-difference (DID) method and using a pilot of subsidies for electrical connections implemented in a number of groups of households in the six D&E regions. The next two sections of the memo describe the planned designs in greater detail. We do not propose any plans for the Hydropower project at this time because its implementation timeline is more uncertain than those of the other two projects due to environmental considerations.

In collaboration with MCC and their partner in Tanzania, the Millennium Challenge Account-Tanzania (MCA-T), we have written terms of reference (TOR) for a data collection firm to conduct baseline surveys of households, enterprises, and communities. MCA-T will soon issue a request for proposals (RFP) based on that TOR.

We have developed draft sample sizes for this study based on rough approximations that were greatly limited by available information. We recommend that these sample sizes be updated once additional information becomes available. In particular we believe that they could be greatly improved using data from the national census, last conducted in 2002, and the current version of the National Master Sample (NMS) of primary sampling units (groups of households used for conducting surveys in Tanzania). We used the 2001 National Household Budget Survey (HBS) for our preliminary power calculations. These could also be somewhat enhanced using the most recent HBS, conducted in 2007.

In designing these studies, we are striving for a balance between rigor, policy relevance of the evaluation questions, and political feasibility. Thus, we will consult our project officer, MCC staff in Tanzania, and energy sector experts throughout the research process. Although we hope to conduct multiple studies, budget limitations will likely preclude conducting all of the studies discussed here. Even so, MPR plans to evaluate both the Cable and D&E projects, looking at both business and household impacts for the D&E project.

[1] The term transmission and distribution (T&D) is also sometimes used to describe this set of projects, though that term can also be used to describe a larger set, including some of the work in Malagarasi.

TABLE 1

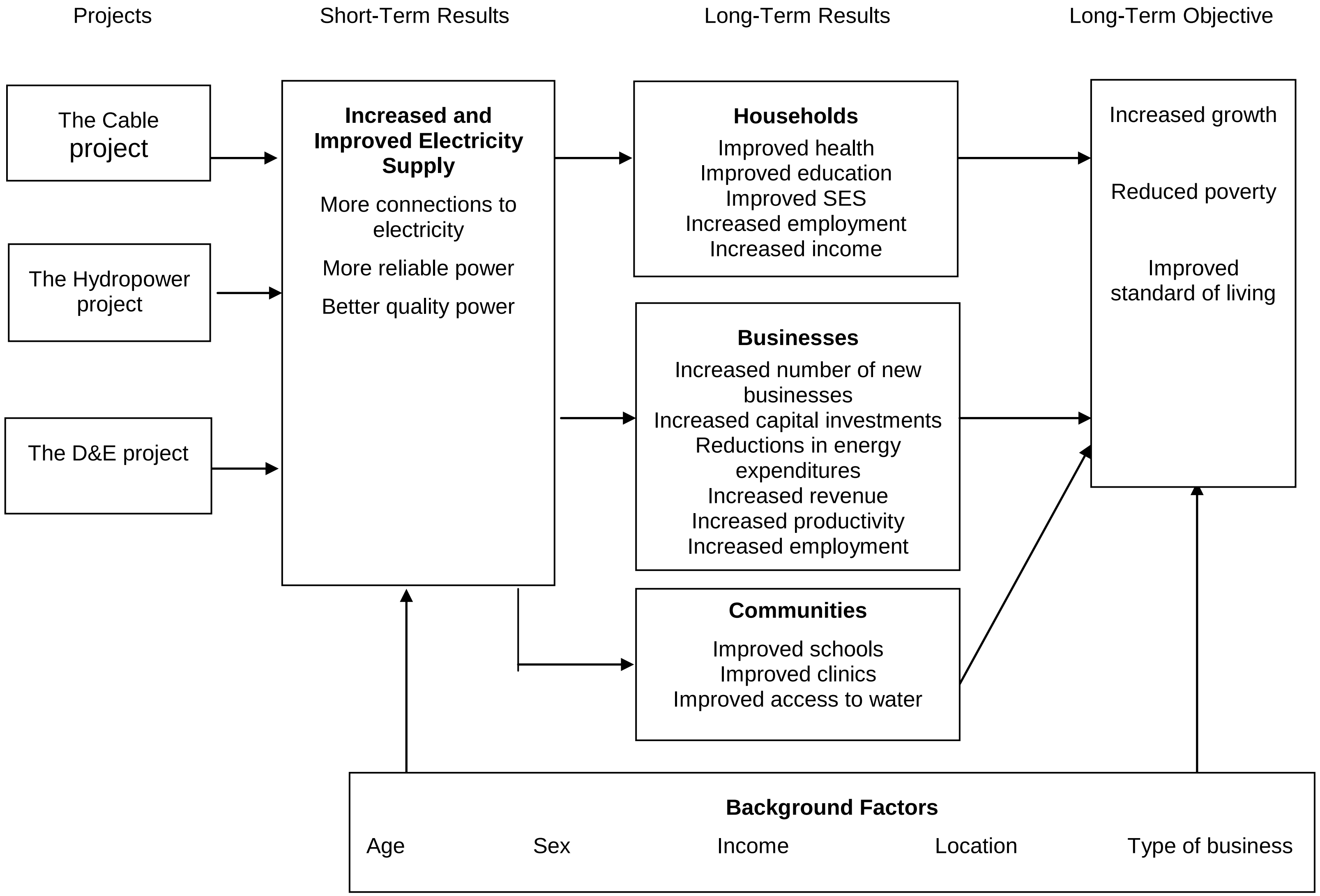
SUMMARY OF STRATEGIES, COMPARISON GROUPS AND VARIABLES OF INTEREST (ENERGY SECTOR PROJECT)

Project/Activity	Evaluation Questions	Proposed Methodology	Treatment Group	Comparison Group	Variables/Outcomes of Interest
Zanzibar Interconnector Project/ Cable and Rehabilitation Activities (Cable project)	Does the project lead to improved reliability and quality of electricity?	ITS	Months after activity is finished	Months before activity begins	Power reliability and quality Household income and expenditures Business costs Employment Health outcomes Child schooling attainment (or at least intensity of study) Intra-household contribution to income and allocation of resources
	Does increased reliability and quality of electricity lead to reduced energy-related costs for hotels?	PPC	Hotels after activity is finished	Hotels before activity is finished	
Malagarasi Hydropower Project / Dam and Extension Activities (Hydropower project)	No evaluation is currently planned for this intervention.				
Distribution Systems Rehabilitation and Extension Activities (D&E project)	Does access to electricity lead to increased household income and better health and education outcomes?	RD or DID	Households in subprojects that received increased access due to nearby line extensions	Households in subprojects that do not receive increased access, due to a lack of line extensions	
		SP	Households in clusters that receive increased access due to connection subsidies	Households in clusters that do not receive increased access due to a lack of connection subsidies.	
	Does access to electricity lead to increased business activity, including new firms, capital investments, and greater levels of employment?	DID	Businesses near completed line extensions	Businesses before lines are extended	

*Access to electricity refers to being connected to a new line funded by MCC, receiving a subsidy to connect to one of these new lines, or actually being connected. Quality of electricity refers to voltage fluctuations while reliability refers to interruptions in the power supply. ITS=Interrupted Time Series, PPC=Pre-Post Cost Study, RD=Regression Discontinuity, and SP=Subsidy Pilot.

FIGURE 1

CONCEPTUAL FRAMEWORK FOR THE ENERGY SECTOR PROJECT



B. THE CABLE PROJECT

The Cable study is designed to answer two questions related to the MCC-funded cable project in Tanzania.

- Does the project lead to improved reliability and quality of electricity and to what extent?
- Does increased reliability and quality of electricity lead to reduced energy-related costs for hotels and to what extent?

To answer the first question we will collect monthly data from ZECO on outcomes such as the number of power disruptions, number of customers, KWH sales, KVA sales, energy/peak demand billed by month, and sales for several months before and after the project has been completed. It may be possible to get data by customer type, but data may only be available for Zanzibar as a whole. We will compare the data by month from before and after project completion using the ITS method. The Cable project can affect the reliability and quality of electricity because the current cable that brings electric power from the mainland to Zanzibar does not have sufficient capacity to serve the current demand for electricity in Zanzibar. Consequently, electricity there is often unreliable and of low quality.

To answer the second question we will use a PPC method. This would involve collecting cost data from a sample of hotels in Zanzibar for two points in time: before and after the project is completed. Improved electricity reliability and quality can affect hotel costs because hotels are often forced to use higher-cost energy sources when electricity is not available and to pay for repairs of appliances that have been damaged by power surges. We are focusing on hotels because they are one of the major businesses in Zanzibar and so that we can design a survey that is well targeted toward their uses of electricity.

One important project implementation issue is that distribution problems within Zanzibar may limit the benefits from the upgrade of the cable. This could minimize the impacts we detect as a result of this study.

C. THE D&E PROJECT

There are two questions that we will answer related to the D&E project.

The first question is "Does access to electricity lead to increased household income and socioeconomic status (SES)[2] and better health and educational outcomes and to what extent?" We are considering several different methods to investigate this question.

Our original plan was to use the RD or "ordered list" method to estimate impacts of extending lines to new areas in the six D&E regions. To use this method, we would need a list of potential line extension subprojects. The determination of which subprojects are implemented must be based almost entirely on one variable, such as the revenue/capital expenditure ratio. To illustrate the method this list could be sorted in the order of that variable. In order for the method to work almost all potential subprojects with values above some cut-point receive services and almost all of the others do not. During a trip to Tanzania in February 2009, we discovered two major problems with using RD for the D&E study. The first issue is that funding decisions were not solely based on the revenue/capital expenditure ratios. Rather, other factors played a major role. The second is that the sample size of unfunded subprojects below the cut-point appears to be very small.

For these reasons, we are also considering using a DID method to study the D&E project. To implement the DID method, we would compare changes in outcomes for treatment areas with changes in outcomes for comparison areas. The treatment and comparison areas would be chosen to be similar based on observed characteristics. The changes would be measured during the period when electricity lines were extended in the treatment areas but before lines were extended in the comparison areas. Thus, the baseline survey would be conducted before projects were conducted in either set of areas while the follow-up survey would be conducted after projects were completed in the treatment areas but before projects were completed in the comparison areas. The comparison areas could be chosen using an "early and late" (DID-EL) approach in which case the comparison areas would be chosen from subprojects that were going to be completed after the follow-up survey while treatment areas would be chosen from subprojects that were going to be completed well before the follow-up survey. The treatment and comparison areas could then be matched based on their revenue/capital expenditure ratios.

[2] SES is used here to refer to other measures of household well-being based on factors such as consumption and ownership of durable goods.

We suspect that we will need to wait at least one year after lines are put in place and perhaps longer in order to see significant benefits of line extensions. First, it may take a few months for households to pay for electricity, another month for the electricity to be installed, another month for them to purchase appliances to take advantage of the electricity and a few more months to learn how to use those appliances to full advantage. Many of these steps may take longer so the net result is that one year may not even be long enough to see impacts.

This means that we would have to wait at least one year after the treatment subprojects are completed before conducting the follow-up survey in order to capture the full benefits of increased access. Any subprojects that are completed during this time period could not be included as comparison projects (assuming a fairly short period for the follow-up survey). This means that we may not have enough "early" and "late" projects to use the DID-EL option. If this happens then we might add some additional comparison groups of households from non-subproject areas to our study and conduct a "no-treatment comparison group" (DID-CG) method meaning that many of the comparison areas would have no electricity projects planned at all. To use this method we would have to match the treatment and comparison group areas on something other than the revenue/capital expenditure ratio since that would only be available for the subproject areas. Data from the 2002 Census might be useful for this purpose.

During our visit to Tanzania, we met with funders from a number of organizations, such as the World Bank, The Norwegian Agency for Development Cooperation (NORAD), and Swedish International Development Cooperation Agency (SIDA). These donors expressed strong interest in knowing more about the impacts of subsidies for electrical connections. In a recent pilot project, SIDA subsidized 85 percent of the costs of electrical connections for almost 1,000 customers. We suspect that some other donors, like the Global Partnership for Output-Based Aid (GPOBA), might also be willing to subsidize connection costs if they find that such subsidies improve well-being. Indeed, GPOBA has already provided such subsidies in Armenia.

Consequently, we are considering using the "subsidy pilot" (SP) method to learn more about how access to electricity affects household income, SES status, and health and educational outcomes. This would involve providing subsidies to cover some part of the fee to connect to the grid to a pilot set of households that are near to the new line extensions and then comparing their outcomes to those for similar households without such subsidies and to households without new lines. This would enable us to estimate impacts of two types of access to electricity—increased access caused by subsidies and improved access caused by the combination of subsidies and new lines. The first set of estimates could be obtained by comparing outcomes for households with subsidies and lines to households with lines but no subsidies. The second set of estimates could be obtained by comparing outcomes for households with subsidies and lines to households without lines.

Our current plan is to use the SP method in the six D&E regions. To implement this method, we would choose a diverse set of household clusters, each of which would be eligible for these subsidies, and two similar sets of comparison households—one with lines and one without lines.

The RD and DID methods both measure some of the potential benefits of electrifying clinics, schools, water sources, and other community infrastructure, as well as the benefits of having electricity at home and of living in a neighborhood where others have electricity. The SP method focuses more on measuring the benefits of having electricity in one's home and in the neighborhood.

The second question related to the D&E study is "Does access to electricity lead to increased business activity, including new firms, capital investments, and greater levels of employment?" We plan to use one of the DID methods and data collected through an enterprise survey to estimate the impacts of new line extensions on business activity.

Delays in project implementation could affect our study plans. During a trip to Tanzania in February 2009, several of the people we met with noted that the major provider of electricity in Tanzania, TANESCO, is probably not recovering all of its costs. This may lead to delays in connections to the extended line, which could in turn delay when we would expect to see impacts of the D&E project on household and business-level outcomes. To help deal with these delays we have developed a flexible design option that would enable us to determine which evaluation method we use for the D&E project after the baseline survey is conducted. This flexible design option is described in Appendix B.

APPENDIX A

TECHNICAL DETAILS FOR MPR TANZANIA ENERGY SECTOR DESIGN REPORT

This appendix expands on our ideas for proposed studies of the MCC-funded electricity projects in Tanzania. A later, design memo will be needed to finalize these plans.

A. GENERAL DATA COLLECTION ISSUES

We plan to collect survey data from households and businesses both before the treatment group subprojects are implemented (at baseline) and afterwards (at follow up). We will also collect baseline and follow-up data from a comparison group which will be chosen so that no line extensions are implemented there until after the follow-up survey. We have written a TOR for a data collection firm. MCA-T is issuing a request for proposals (RFP) based on that TOR. The TOR covers baseline data collection for household and community surveys and also for the enterprise survey that will be conducted in one region in the mainland. It does not cover the follow-up surveys or the data collection in Zanzibar.

Before collecting the baseline data, we will need to determine how many subprojects, clusters of households per subproject, and households per cluster to use. We have included a preliminary set of numbers for the TOR but recommend that these estimates be updated based on additional information regarding the actual numbers of households per PSU likely to be eligible for our study. Further information on the correlations between outcomes within PSUs would be helpful and could be estimated using 2007 HBS data (the current estimates are based on 2001 data). Estimated costs of collecting data from additional PSUs and estimated costs of collecting data on additional households within PSUs could also be helpful and was requested as part of the TOR.

We are still developing a method for selecting clusters of households. We will need to develop a method that does not overburden either ourselves or the local data collection firm. We anticipate that it will be necessary to have staff from the local research firm walk along large portions of the expected locations of the lines in each subproject area and identify clusters of households to be used for the sample frame for our study. This is how the TOR was written. We believe, however, that we could significantly reduce the miles of walking required if we could obtain information on the exact locations of the lines and combine that with data from the NMS of primary sampling units (PSUs) so that we could identify which parts of the lines are most likely to provide useful clusters of households for our study. This information would be required at least a few months before the survey firm staff walked the lines so that we could process and analyze the resulting data. This possibility is discussed further in the section on power below. Once we have developed a method of selecting clusters of households, we would then select a sample of households in each cluster.

B. CABLE PROJECT STUDY

We plan to investigate two questions for the Cable project study. The first is “Does the project lead to improved reliability and quality of electricity and to what extent?” To answer this question we propose to use the ITS method. The second question is “Does increased reliability and quality of electricity lead to reduced energy-related costs for hotels and to what extent?” To investigate this question we propose using the PPC method.

General Cable Project Implementation Issues

The work we propose to do on the Cable project could be constrained for two reasons. First, if the Cable project is delayed until too near to the end of the study, it may not be possible for us to conduct research on it using either the ITS or PPC methods because there would not be enough time for us to collect and analyze the follow-up data before the end of the study. Second, electricity distribution problems within Zanzibar may limit the benefits from the upgrade of the cable. If we find no impacts we will probably ask engineers for the Zanzibar electricity distribution firm (ZECO) whether or not this was a likely issue. This would suggest that the lack of impacts may have been caused by problems in other parts of the electricity grid rather than problems with the new cable itself.

Question 1: Does the project lead to improved reliability and quality of electricity?

Method: ITS

To answer this question, we propose to analyze ZECO data on the number of customers, KWH sales, KVA sales, energy/peak demand billed by month, and sales by customer group. These data are reported monthly for the island of Zanzibar. We might also be able to obtain data by customer type. This would help to identify where impacts of the cable might be largest. However we were warned that the breakdowns by customer type might not be reliable.

The ITS method will enable us to specifically answer two questions: (1) How does electricity reliability and quality change after this project is completed? and (2) Do these changes appear to be significantly different from changes observed during similar periods of time before the project was completed?

Several of the people we met in Zanzibar complained that a lack of electricity reliability is a major problem there, so this would be a focus of our Cable study.

Description of ITS Method

The ITS method involves analyzing how the outcomes described above change on a monthly basis for a number of months before and a number of months after the project is implemented. Using these data we would conduct a statistical test to see whether or not the changes observed when the project is implemented differ significantly from those observed during other periods of time of similar length. Evidence that these outcomes improved when the project was implemented in ways that differed statistically from changes during other periods would suggest that something unusual did occur at that time, which would be consistent with the project having an impact.

ITS Design Issue

This method would not enable us to definitely conclude that the Cable project caused the observed changes; we could only determine whether or not the observed change was unusually large (and positive) and thus consistent with a causal effect. It might be caused by the MCC-funded project but could also be caused by other changes that occur at the same time such as improvements in the grid in Zanzibar paid for by non-MCC funds.

ITS Data Collection Issues

We do not expect to have any difficulty collecting these data but we have not yet seen sufficient data to be sure that they will suffice for our purposes. Also, it is possible that ZECO and/or TANESCO will change how they are collecting these data shortly before or during the project period in a way that makes it impossible to make comparisons over time. This seems unlikely to happen for all outcomes, but may happen for some subset of the outcomes described above.

Question 2: Does increased reliability and quality of electricity lead to reduced energy-related costs for hotels?

Method: PPC

Using the PPC method, we would investigate whether the upgrade of the cable leads to reduced costs of operation per customer for hotels in Zanzibar. We could also look at the impact on other outcomes. We propose focusing this study on hotels, since tourism is one of Zanzibar's major industries.

In theory there would be no benefit to looking at this question if the answer to the first question turns out to be that the Cable project did not improve the reliability and quality of electricity. Unfortunately, as explained in Appendix B, we may not have enough time to answer the first question before we need to collect data for the second question. Nonetheless, when we write our final report we will certainly incorporate information on the answer to the first question when discussing the second.

The PPC method enables us to answer the following sub-questions related to Question 2 above:

- What changes were observed in energy-related costs such as reduced repair costs for electrical appliances after the project was implemented?
- Are those changes plausibly related to the project?

Description of PPC Method

To implement the PPC method we would collect cost data from a sample of hotels in Zanzibar before the upgrade of the cable. This cost data might cover outcomes such as the amount of money spent repairing electrical appliances like TVs and air conditioners in a month. It will be important to stratify on the size of the hotel, since small hotels are more vulnerable to power shortages, as they are less likely to have generators. We would then return 12 months after the new cable is active and collect the same data. It might be better if we could wait longer to return, say after 24 months, so that more data would be available, but that might not be possible given the time frame we have for the study.

The PPC and ITS methods are similar in that both are designed to see if outcomes changed after the project was implemented. However, they differ in the sense that the ITS method looks at variation over time (by month in our case) to measure uncertainty whereas the PPC method uses variation across observations (hotels in our case) to measure uncertainty. Thus, if we find a statistically significant result using the ITS method this means that the change observed is larger than would be observed if we were to pick two random sets of months with similar sample sizes before or after the project was implemented and compared their outcomes. In contrast, a statistically significant finding using the PPC method indicates that the difference is larger than what would be observed if we were to pick two random sets of hotels with similar sample sizes at either point in time and compared their outcomes.

PPC Design Issues

As with the ITS method, we would not be able to determine with much certainty whether or not the changes observed when the project was implemented were caused by the project or by some other factors that changed at the same time. In addition using the PPC method we would not be able to determine whether or not the changes observed were unusual given the amount of time that had passed, something that is possible using the ITS method. To help address this issue we will include open-ended qualitative questions on the survey asking the hotel owners whether or not they believe that electricity quality or reliability may have changed due to non-MCC funded work.

PPC Data Collection Issues

While a number of businesses have indicated that they have the types of data we are interested in collecting and that they would be willing to provide us with these types of data, we are still not sure that they actually will do so. In addition, we will have no way of knowing whether or not their data are accurate. However, we will promise the businesses confidentiality especially for the more sensitive data on costs. Also, a great deal of the data, for example on the types of appliances purchased, may not be very sensitive.

C. D&E PROJECT

There are two questions that we will answer in our D&E project study, one related to households and a second related to businesses.

1. Does access to electricity lead to increased household income and socioeconomic status (SES), and better health and educational outcomes and to what extent?
2. Does access to electricity lead to increased business activity, including new firms, capital investments, and greater levels of employment and to what extent?

We are considering several different methods to investigate the household question. These include the RD, DID, and SP methods. The RD and DID methods both capture some of the potential benefits of electrifying health clinics, schools, water sources, and other community infrastructure, as well as the benefits of having electricity at home and living in a neighborhood where others have electricity.

To answer the second question, which focuses on businesses, we plan to rely on data collected from our enterprise survey.

General D&E Project Implementation Issues

The D&E project includes subprojects that cover both extension of electrical lines into new areas and rehabilitation of existing lines. We are currently planning to focus on the extension subprojects rather than on the rehabilitation subprojects since there are few subprojects that focus on rehabilitation. The timing of the subprojects will affect whether or not we can estimate impacts of the line extensions, as discussed below.

Question 1: (Households): Does access to electricity lead to increased household income, SES, and better health and education outcomes?

Method 1: RD (the "ordered list") approach

Our original plan was to use the RD method to estimate impacts of extending lines to new areas in the six D&E regions. To implement RD, we would need to start with a list of potential recipients of new services, only some of whom will actually receive services. The list must be organized based on a variable that was used to determine which potential recipients would receive the services. The determination of who receives services must be based almost entirely on that variable, such that almost all potential recipients with values above (or below) some cut-point receive services and almost all of the others do not. The estimated impacts of extending the lines will be based in large part on comparisons of outcomes for recipients just above and just below the cut-point.

For this project, the potential recipients are households in subprojects. The subprojects were all designed by TANESCO and include some that MCC is planning to fund and some that MCC is not funding. The variable used to determine eligibility for MCC funding for services is the estimated revenue/capital expenditure ratio for the subprojects.

The RD method (and the DID method discussed below) will specifically enable us to answer the following questions.

1. What impact does extending the lines have on household connection rates?
2. What impact does extending the lines have on businesses connection rates?
3. What impacts does extending the lines have on access to schools and health facilities with electricity?
4. What impacts does extending the lines have on water and road projects?
5. What impacts does extending the lines have on well-being (income, SES, health, education)?

Extending the lines may cause local governments to also improve water systems and roads. Thus, the impacts of extending the lines on households may be both direct, because of increased access to electricity, and indirect, because governments choose to invest more money in water and road projects in areas with electricity.

RD Design Issues

There are two major problems with using RD for the D&E project, and a few minor problems.

Major Issues with RD Design. The first major issue with using RD to study the D&E project is that funding decisions were not solely based on the revenue/capital expenditure ratios. Rather, other factors played a major role. The second major issue is that the sample size of comparison subprojects appears to be very small, given the current cut-point.

We learned about these issues during a meeting Duncan Chaplin had with Isaac A. Change, director of energy sector projects, MCA-T and others at MCA-T on the last day of his visit to Tanzania in February 2009. This meeting was arranged to obtain information on how MCA-T chose the projects to be funded using MCC monies. MCA-T told Dr. Chaplin that they used an indicator (the revenue/capital expenditures ratio provided by the firm Hatch Mott McDonald (HMM)) to select the subprojects to be funded. In general, subprojects with ratios above some cut-point were chosen for funding. MCA-T did not tell Dr. Chaplin what the cut-point was, but he was shown a list of subprojects along with their ratios, and MCA-T identified which of them were to be funded by MCC and which were to be funded by other donors. There was no clear cut-point dividing the funded projects from those not funded. For the rest of this memo we assume a cut-point of 0.034 (similar to that suggested by our project officer, Anne Rothbaum). Mr. Chanji explained that the cut-point was not clear for the following reasons:

First, both TANESCO and other funders chose a few subprojects that had ratios below 0.034. TANESCO chose five such projects for MCC funding, and other funders chose two more. Second, at least one subproject above the cut-point may have been dropped because it was too large, given available funding. Thus, other subprojects may have been chosen in part because they were smaller, even though they had slightly lower ratios.

These two explanations do not appear to explain all of the discrepancies. In checking the list, Dr. Chaplin noticed that there were many subprojects with ratios above the 0.034 cut-point that were not funded. Mr. Chanji was not able to explain why this happened.

Following is a summary of the list that was later provided to MCC by Mr. Chanji via Chedaiwe S. Luhindi, Monitoring and Evaluation Officer, MCA-T.

- There are 240 subprojects with income/capital expenditures ratios.
- For those with ratios over 0.034, the probability of being funded was 90.5 percent (199 out of 220).
- For those with ratios below 0.034, the probability of being funded was 35 percent (7 out of 20).

As a result, there are only 13 subprojects that were both below the cut-off and not funded. These sample sizes will likely result in very imprecise estimates if we use an RD design to estimate impacts. A later list showed that five of the subprojects above the line that were not funded according to the original list were then expected to be funded. That increases the probability of funding for projects above the line to 92.7 percent.

Minor Issues with RD Design. There are also three minor issues with the RD design. These relate to mobility, confounding, and contamination.

Mobility could bias our estimates if the introduction of electricity to an area is associated with mobility changes that result in a change in the average income level. This would be particularly problematic if, for example, poor people were less likely to move out of treatment group areas than comparison group areas because of the additional access to electricity in the treatment group areas. Our site visits suggested that there may be substantial household mobility when electricity is introduced. Some people may move into an area with new electrical lines to take advantage of the electrical connections, and others may leave. This could greatly complicate our ability to estimate impacts. We can track how many of our baseline respondents move, and ask about migration in our follow-up study. If we find that out-mobility is common, we would need to acknowledge this as a potential source of bias in our results. Dinkelman (2008) did find evidence that electrification impacts mobility, though in her study it was not sufficient to have a major impact on her estimates of the benefits of electrification. This mobility could also impact our DID and SP method estimates.

Confounding could bias our estimates if, in some locations, multiple government, MCC, or other donor-funded water and road interventions are implemented during the time the MCC-funded electricity projects are being implemented. In those cases, we will be estimating the impact of these combined interventions. We will make sure to obtain information regarding concurrent interventions during our baseline and follow-up surveys. This could also matter for our DID method.

Contamination could cause us to underestimate the benefits of electrifying schools, health clinics, water suppliers, and other community facilities if some people receive some of these community services from areas outside of the subproject areas we are studying. While both the RD and DID methods will enable us to estimate some of these benefits, we do not expect our estimated impacts to incorporate the full benefits of these shared facilities. We are currently investigating methods that might enable us to capture more complete benefits of electrifying communities. This work may be facilitated by the data we collect in our community survey.

RD Data Collection Issues

Sample Sizes. For the RD method we plan to use subprojects from all six regions of Tanzania, if economically feasible. However, we still do not know (1) how many subprojects above and below the line we will use and (2) how many households we will need per subproject. Preliminary power calculations are presented in Appendix B to help inform the determination of those sample sizes.

Identifying Households. We are currently considering the importance of whether households are located at the beginning, middle, or ends of the extended lines. If we want to focus on households that have higher-quality electricity, we would focus on those at the beginning of the lines. If we do this, we would also need to select households in comparison sites (that is, the unfunded subprojects) that were likely to have been at the beginning of an extended line were a line constructed there.

Intervention Households. In February 2009 MPR staff met with Francois Pienaar, QC and QA specialist, of the firm ESB International (ESBI). This is the firm that is designing the electricity projects being funded by MCC in Tanzania. Mr. Pienaar said that ESBI may be able to provide MPR with GPS coordinates, maps, or both, so that we know exactly where the lines will be in the treatment sites. Alternatively, we can get some information about these locations through the tenders that will be issued by ESBI to the firms that will be doing the electrical rehabilitation and extension work funded by MCC.

After obtaining information on the exact locations of the new lines, the next step would be to have the local research firm go to the areas and create a sampling frame of households that meet certain criteria. These criteria might include having no mud walls or grass-thatched roofs and being within 30 meters of where the lines will be extended. The 30-meter rule would be used because TANESCO (the electricity firm) provides incentives to encourage its staff to quickly connect all households that pay their connection fees and are within 30 meters of the line.

With oversight from MPR, the local research firm would select households at random from the sampling frame for each subproject to be included in the analysis.

Comparison Households. In order to use the RD method, we will also need to select comparison households in areas that did not get new electrical lines but where TANESCO does have plans for future subprojects involving electrical line extensions. The key challenge here is that we do not know exactly where ESBI would have chosen to put these lines had MCC decided to fund these subprojects. Thus, we do not know which households to survey. Both Denzel Hankinson and Mr. Pienaar have weighed in on this issue. It appears that there are a few ways we can address this problem:

- ***Gold Standard Approach.*** Hire ESBI to determine where the lines would be in the unfunded subprojects. We have been advised by our project officer that this is not possible, since ESBI is very expensive and MCC does not want to enter into new contract negotiations with them.
- ***Weakest Approach.*** Send the data collection firm to the general area where the subproject is and have them choose households that appear to be in areas that would get new lines—in particular, areas with lots of households that have tin roofs, do not have mud walls, and are near major roads. This is a risky approach, as it may require the local research firm to make some difficult judgment calls. The firm's staff members are unlikely to have engineering experience, so they will not be able to determine whether landscape steepness, vegetation, buildings, or other features would make it impossible to extend the line in a certain location.
- ***Middle-of-the-Road Approach.*** Hire a local engineering consultant (a retired TANESCO staff member or even one of the people ESBI used) to work with the local data collection firm to determine where the lines would have been extended.

With the information we have now, we think the third option is most appropriate, but we need to have further discussions with our project officer and gather more information before making a decision. After determining where the lines probably would have been, the research firm would follow the same procedure as for the intervention households to select individual households for our household survey. This is what we asked for in the TOR for data collection.

RD Project Implementation Issue

Currently we do not expect to be able to use the RD method due to an insufficient sample size of subprojects in our comparison group and because it appears that funding decisions were affected greatly by factors other than the revenue/capital expenditure ratio. This could change, however, if the number of subprojects that MCC can fund is reduced significantly (either because of a drop in funding or because costs of project implementation increase) and if the updated list of funded project is based more on the revenue/capital expenditure ratio.

Method 2: DID ("early-late" or "no-treatment comparison group")

Description of DID Method

The DID method would involve selecting two sets of households—one set that is expected to receive their electricity extensions before we collect our follow-up data (the treatment group) and one that is expected to receive their intervention later or not at all (the comparison group). We would then collect baseline data from both groups before the treatment group receives electricity. Later we would collect follow-up data from both groups considerably after the treatment group receives electricity but before the comparison group receives electricity.

We could select the comparison group in one of two ways. First, we could select clusters of households in areas that are scheduled to receive their electrical connections during the study period, but after our follow-up data collection. This can be thought of as the "early-late" (or "phased-construction") method, as the treatment group receives their interventions early while the comparison group gets their interventions late. If we used this method, we might try to match the treatment and comparison groups based on the revenue/capital expenditure ratios of their subprojects. The other way to select a comparison group would be to select clusters of households that are not expected to get electrical connections at all during the MCC program period either because they are too far down on the list or because they are not on the list at all. This can be thought of as the "no-intervention comparison group" method. This would mean that we would probably not have revenue/capital expenditure ratio data for at least part of the comparison group. Consequently, if we used this method we would probably try to match the comparison and treatment group clusters of households based on other characteristics—perhaps using data from the Tanzanian census. If we are able to obtain data on a sufficient number of characteristics for a large group of clusters of households then we could use propensity score matching (PSM) methods to choose the treatment and comparison groups. PSM is a statistical method of matching groups of households based on multiple criteria. To identify households in these clusters we will look for roads that are similar to those where electricity is made available in the treatment clusters and identify households along those roads.

DID Design Issues

The DID method gets rid of one major issue associated with the RD method—that of having too small a sample size of comparison group projects.

However, the DID method also suffers from the three minor issues associated with the RD method (mobility, confounding, and contamination). Nevertheless, compared to the SP method, we would expect the DID method (like the RD method) to be able to capture more of the benefits of electrifying schools, health clinics, and water suppliers.

The DID "early-late" method has two benefits compared to the DID-CG method. First, when using the DID-EL method we will have accurate information on the location of the lines for both the treatment and comparison groups (This is a major benefit over the RD method as well). Second, when using the DID-EL method we can match the treatment and comparison groups based on their revenue/capital expenditure ratios in addition to any other data we obtain from other sources (perhaps the 2002 Census). In contrast, with the DID-CG method we can only use other data to do the matching. Thus, the chances that the treatment households and comparison households have unobserved differences that influence trends in their outcomes is somewhat larger using the DID-CG method than using the DID-EL method.

While the DID "early-late" method has some benefits it also has some weaknesses. In particular, the DID "early-late" method depends critically on the timing of the projects, the timing of the availability of information on the project timelines, and the accuracy of that information. If the subprojects are all conducted within a relatively short time period, we will not be able to use the DID "early-late" method. Similarly, if we do not have an accurate description of when subprojects will be done, we will likely end up conducting surveys in the wrong areas, thus compromising the potential of the DID "early-late" method. In both of those situations the DID "no-intervention comparison group" method would be less risky.

Method 3: SP method

Description of SP Method

During our visit to Tanzania in February 2009, we met with funders from a number of organizations, such as the World Bank, NORAD, and SIDA. These donors expressed strong interest in knowing more about the impacts of subsidies for electrical connections. Indeed, in a recent pilot project, SIDA subsidized 85 percent of the costs of electrical connections for almost 1,000 customers. Similarly, GPOBA has provided subsidies for electricity in Armenia. We suspect that other donors might also be willing to subsidize connection costs if they find that increasing access to electricity improves well-being.

For this reason we are considering estimating the impacts of subsidies (and subsidies with line extensions) by subsidizing electrical connections in a carefully chosen set of households in the six D&E regions and then comparing those households to similar households in the same areas that are not eligible for subsidies and to similar households in areas without lines. We refer to this as the SP method.

The SP method will enable us to answer the following questions:

1. What impacts do subsidies and subsidies with line extensions have on connection rates in the six D&E subproject areas?
2. What impacts do these subsidies and subsidies with line extensions have on well-being (income, SES, health, education in these areas?

3. Does having an electricity connection lead to increased household income, SES, and better health and education outcomes and to what extent?

To answer the last question we will try to combine the answers to the first two questions. As explained below this may be difficult.

To see how we plan to estimate the impacts of being connected it is helpful to think about three groups of respondents—those that would get electrical connections regardless of whether or not they got subsidies (the "Always-Connect" households), those that would never get connected even with a subsidy (the "Never-Connect" households) and those that would connect if and only if they get a subsidy (the "Subsidy-Connect" households). Since being in the treatment group does not affect electricity use for the Always-Connect and Never-Connect households we cannot estimate the impacts of electricity for them. Rather, our estimates apply only to the Subsidy-Connect households. Now suppose that we find that 30 percent of the comparison group is connected and 20 percent of the treatment group is not connected at the time of the follow-up survey. If the treatment and comparison groups are well chosen then we would expect the fractions of "Always-Connect" and "Never-Connect" households to be similar for the treatment and control groups. This would mean that 30 percent of our total sample (treatment and control combined) is "Always-Connect" and 20 percent of our sample is "Never-Connect." This leaves 50 percent for the "Subsidy-Connect" households. Thus, we can only estimate the impacts of being connected for this 50 percent. Now suppose we find that the school enrollment rate for children in the treatment group is 60 percent compared to 45 percent for the control group. This suggests a fifteen percentage point difference. If this difference was caused by only half of the sample then we would expect that the effect must have been about twice as large (around 30 percentage points) for the "Subsidy-Connect" group whose electricity use was actually affected by being in the treatment group.

The SP method will work well for answering the third question as long as the "Always-connect" and "Never-Connect" households remain unaffected by being in the treatment group. This seems like a plausible assumption for the "Never-Connect" group. It seems less plausible for the "Always-Connect" group. For those households the subsidy is a fairly large financial benefit that does not affect electricity connection rates but could easily affect other outcomes. To see this note that a $100 subsidy is on the low end of what we are thinking of offering but is still around 20 percent of average annual Gross Domestic Product per person in Tanzania in 2009.[3] This means that it could easily have large impacts on other outcomes we care about through pathways other than access to electricity.

One way to mitigate this problem is to select the sample in a way designed to exclude the "Always-Connect" households. This might be done by using a means test perhaps based on the

[3]http://www.imf.org/external/pubs/ft/weo/2009/01/weodata/download.aspx.

quality of their home (building material and number of rooms) and the consumer goods they owned (car, TV, etc.). In addition, we could wait for a certain period of time (say six months) after the lines were installed before offering subsidies in the hopes that the wealthier households would connect before the end of that period. Finally, once we have conducted the follow-up survey we could use data collected in the baseline survey from the comparison group to determine which characteristics predict which households are "Always-Connect" (that is households that are connected at the time of the follow-up survey when in the comparison group). We can then use this information to exclude households with these characteristics from both the treatment and control group when doing our analyses. In combination these methods should help to mitigate the possibility that our results are affected to a large degree by the income benefits of the subsidy obtained by the "Always-Connect" group.

To ensure good answers to all three types of questions we will select the treatment and comparison groups to be as similar as possible. For the subsidized households and unsubsizied households with lines this will happen after conducting the baseline survey to ensure that the interviewers are unaware of the treatment group status of the clusters when they collect the baseline data and cannot, therefore, unintentionally alter how they collect the data based on that knowledge. The groups of households without lines will probably be selected before conducting the baseline survey so that we do not need to survey additional groups of households.

SP Design Issues

Our current plan for the SP method is to pilot subsidies for electrical connections for households in the six D&E regions. For this pilot, we would choose a diverse set of clusters of households, each of which would be eligible for these subsidies. The clusters chosen will likely be villages or subsets of villages and may not map directly onto existing administrative units.

We would select households within a cluster that are geographically close to each other. This would enable us to estimate some of the indirect impacts of having neighbors with electricity as well as the direct impacts of having electricity in one's own home. Clusters would be chosen randomly from a larger set of clusters so as to be broadly representative of the six D&E regions. A similar group of clusters would be chosen for a comparison group so that we could estimate the impacts of these subsidies by comparing outcomes for these two groups.

One of the concerns raised by MCC staff in Tanzania is that there might be some potential political problems if treatment and comparison group households are near to each other: The comparison households might become upset that they are not receiving electricity subsidies. Consequently, we will try to choose our comparison group clusters of households so that they are far from the treatment group clusters, though still within the same subprojects to help ensure comparability.

It is unclear how many of the D&E subprojects will be funded. The current plan is to fund all of the subprojects above the cut-point, but if funding is limited, this may result in a smaller number of subprojects being funded. We assume that if some subprojects are not funded, those closest to the cut-point will be eliminated first. Consequently, we may select subprojects for our

SP that are far enough above the cut-point line that they are likely to receive funding. We will determine the cut-point to use for the SP based on discussions with MCA-T and MCC regarding the possibility that the number of D&E subprojects that are funded is reduced either due to a reduction in funding or an increase in costs of these projects.

Our understanding is that many low-voltage lines in Tanzania are too long, so the quality of the electricity available degrades noticeably between the beginnings of the lines, where they connect to the higher voltage lines, and the ends. In the long run it is reasonable to expect that the quality of electricity will rise as longer high-voltage lines are created. Consequently, from a policy perspective it may be of greater interest and value to understand the impact of high-quality electricity on household well-being. To focus on the effects of high-quality electricity in the SP, we could choose clusters of households near the beginnings of the low-voltage lines. This issue could apply to the RD and DID methods as well.

Implementing the subsidies will be non-trivial. We will need to make sure that the households are aware of the subsidies, that they have time to pay, and that their connections are made, all in time for us to observe changes in outcomes by the time of the follow-up survey.

To avoid political backlash, we plan to offer subsidies to all households within a given geographic unit (cluster of households). Unfortunately, once electricity is introduced, many new people may move into the area, substantially increasing the number of households in the geographic cluster. If we promise to make all households in a given cluster eligible, it may be impossible to accurately predict how many households will end up being eligible for subsidies. Thus, we would either have to give the subsidies only to the people who were originally living in the cluster or reduce the amount of the subsidy to be able to cover the additional households.

Some households in the treatment group will not make use of the subsidies. This means that the estimated impact of being in the treatment group is probably less than the impact of being connected. For this reason it is tempting to include only the households in the treatment group that make use of their electricity subsidies. This would be a major mistake. In order to obtain unbiased answers we need to make sure that the treatment and comparison groups are similar based on both characteristics we can observe and characteristics that we cannot observe. One important example of a characteristic that we cannot observe is whether the households are in the "Always-Connect", "Subsidy-Connect," or "Never-Connect" sets of households. We can identify the "Always-Connect" households in the comparison group because they are the ones that connect—but we cannot identify the "Always-Connect" households in the treatment group since both they and the "Subsidy-Connect" households connect. Similarly, we can identify the "Never-Connect" households in the treatment group since they do not connect but not in the control group where both they and the "Subsidy-Connect" group remain unconnected. We cannot identify the "Subsidy-Connect" households in either the treatment or comparison groups since in both cases we cannot distinguish them from one of the other two sets of households. Thus, if we want to make sure that the treatment and comparison groups are similar based on this unobserved characteristic (and others) we need to include all of the households originally selected for our study in our analyses. A similar point holds for the RD study as we need to make sure all households in subprojects selected for the study are surveyed regardless of whether or not they get line extensions.

Question 2: (Businesses): Does access to electricity lead to increased business activity, including new firms, capital investments, and greater levels of employment?

Methods: RD, DID, and/or PPC

To answer the second question about businesses in the D&E areas, we will probably use one of the DID methods and the enterprise survey data. The current TOR calls for surveying a small set of businesses in one region of the country with D&E subprojects.

APPENDIX B

FLEXIBLE DESIGN OPTION, CHOOSING METHODS, POWER CALCULATIONS, CRITICAL ASSUMPTIONS, NEXT STEPS AND STUDY TIMELINE

This appendix presents a description of a new flexible design option, a discussion of how we might go about choosing which method we will use to estimate the impact of line extensions for the D&E project, and updates on each of the following: power calculations, critical assumptions, next steps, and study timeline.

A. FLEXIBLE DESIGN OPTION

We have developed a new flexible option for choosing our design for the D&E project after the baseline survey is conducted. This option would require conducting the baseline survey in a larger number of subproject areas than originally planned and also in a set of areas where no subprojects are planned using very few PSUs (perhaps only one or two) per subproject and non-project area. In this way we could collect baseline information that could be sufficient regardless of which design we end up using.

While this new option requires spreading out the baseline survey across a somewhat larger number of PSUs than would be ideal for most of the methods we are considering, it would not necessarily require having the followup survey conducted in those same PSUs. If we used the DID-EL, RD, or SP methods then we would only conduct the follow-up survey in the subprojects that were relevant for those designs. This would likely result in much lower costs for the follow-up survey. Alternatively, if we use the RD or SP methods the extra funding could be used to add additional PSUs as neither of those methods requires having a baseline survey. The baseline data collected in the PSUs included in our analyses would still be valuable to help reduce unexplained variance in the outcomes and thus provide more precise impact estimates.

Once we choose our design some of the baseline data will no longer be useful for either our treatment or comparison groups when estimating impacts of access. However, the additional baseline data would not go to waste. If we end up using a design that does not make use of part of the resulting sample, the information on the excluded part would still be valuable for providing information about the characteristics of households that do not yet have electricity in Tanzania. For example, in order to facilitate the possibility of conducting the DID-CG design we would collect baseline data from a group of unfunded projects and areas with no subprojects planned but that are similar to those areas with subprojects. If we instead end up using the RD method the data from the non-project areas would not be used in our impact analyses. However, those areas would likely be representative of the types of areas that will get electricity in the near future as Tanzania's grid continues to expand and thus provide important background information for our study.

Table B.1 presents illustrates a possible set of sample sizes for the flexible option based on the following assumptions.

- Half of the subprojects are funded (around 100).
- There are two PSUs in each of these subprojects and half of those receive subsidies.
- One third of these projects are done early enough to be in the treatment group for the DID-“early-late” method, one third are late enough to be in the comparison group and the remainder would be left out if we use that method.
- The total number of PSUs is 384 (as stated in the TOR recently submitted to MCC).

TABLE B.1

FLEXIBLE DESIGN OPTION ILLUSTRATIVE SIMPLE SIZES

				Treatment Group Status By Evaluation Method				
				DID			Subsidy with	
Subproject Type	Subsidy	Timing	PSUs	CG	EL	RD	EL	RD or CG
Funded	Yes	Early	34	T	T	T	T	T
		Middle	33	T		T		T
		Late	33	T	C	T		T
Funded	No	Early	34	T	T	T	C	C
		Middle	33	T		T		C
		Late	33	T	C	T		C
Not-Funded	No	All	100	C		C		
No Subproject	No	All	84	C				
Total Treatment PSUs				200	68	200	34	100
Total Comparison PSUs				184	66	100	34	100
Total PSUs for Method				384	134	300	68	200

Note: Shaded areas are for treatment group.
T= Treatment group, C=Comparison group.

The first three columns show sample sizes for estimating the impacts of line extensions. If we use the flexible design option and the DID-CG method we would use all of the PSUs where data was collected at baseline and end up with 200 PSUs in our treatment group and 184 in our comparison group (see the first column of Table B.1). This comparison group includes both subprojects that are not funded and PSUs in areas with no subprojects planned. If we end up using the DID-EL method we would have 68 treatment and 66 comparison group PSUs (second column of Table B.1). If we end up using the RD method we would end up with 200 treatment group PSUs and 100 comparison group PSUs.

The last two columns show sample sizes for estimating the impacts of subsidies. If we use the DID-EL method to estimate the impact of line extensions and only did one follow-up survey then we would conduct the follow-up survey before the late projects are completed. This means that we would end up with only 68 PSUs (34 treatment and 34 comparison) for the SP method. If we used the CG or RD method to estimate impacts of line extensions then we would do the follow-up survey after all of the subprojects are complete. This would give us 200 PSUs (100 treatment and 100 comparison) when using the SP method.

B. CHOOSING METHOD FOR D&E LINE EXTENSIONS

Currently we are considering three methods for estimating the impacts of the D&E line extension work—RD, DID-EL, and DID-CG. At some point we will need to choose which method we will use. If we use the flexible option described above then we could probably wait until around six months before the follow-up survey so that we can choose an appropriate sample for that survey. If we are not using the flexible option then we would need to choose which method we will be using at least six months before the baseline survey so that we can draw an appropriate sample for that survey.

Our decisions between these methods will depend on a large number of factors. Following is a discussion of how we would make these decisions. RD is considered the most rigorous of the methods we are considering so we would probably choose whether or not to do an RD study first. If it appears that the list of unfunded projects is likely to be small (say less than 30) or the use of the list is marginal such that more than 40 percent of the projects are non-compliers (above the line and not funded or below the line and funded) then we will likely recommend against using the RD design. If the list of unfunded projects is large (more than 50) and the non-compliance rate is low (less than 10 percent) then we are likely to recommend in favor of the RD method. If other conditions apply we would recommend additional power calculations to determine the optimal method.

If we decide to not do the RD method then we would be choosing between the DID-EL and DID-CG method. This choice will depend in part on the timing of the subprojects and in part on our ability to match the treatment and comparison groups. If we cannot identify which projects will be early and late when we need to make a decision then we would recommend using the DID-CG method. This choice would avoid the risk since this situation would mean that the DID-EL method might not work at all. If we can identify which projects will be early and late in time and we can match at least 50 pairs of treatment and comparison projects for the DID-EL project based on their revenue/capital expenditure ratios then we would likely recommend the DID-EL method over the DID-CG method. If we can only match fewer than 20 such pairs then we are likely to recommend the DID-CG method. If the numbers are in between 20 and 50 then we are likely to recommend conducting additional power calculations to make this decision.

C. POWER CALCULATIONS

In this section, we describe preliminary power calculations we have made as the basis for beginning a discussion of possible sample sizes for this study. These calculations were done using the Optimal Design statistical software package created by Steve Raudenbush for the Grant Foundation. The calculations are based on a rough set of estimates regarding the fixed costs of conducting surveys in Tanzania, the marginal costs of collecting data from additional households, and the marginal costs of collecting data from additional clusters of households. Once we obtain more precise information on these likely costs, the calculations may need to be updated.

Background for Power Calculations

As discussed in Appendix A, we plan to sample households in clusters. To do the power calculations presented here, we needed to make some assumptions about how these clusters would be chosen, and to estimate the likely amount of variation that would occur between and within clusters of households.

We had been considering using clusters of households that were much smaller than a village. After further consideration, however, we now recommend using the PSU that is used by NBS of Tanzania when conducting their surveys. In rural areas PSUs are generally villages or sub villages. We recommend using PSUs for three reasons. First, using the PSU will eliminate the need to develop an alternative method of selecting clusters of households. Second, it appears that there may be a sufficient number of small PSUs such that we could reasonably afford to subsidize all eligible households in those PSUs and still have a large number of PSUs in our study based on the subsidy method.

We had been considering using Google Maps to help select clusters of households but now believe that method probably will not work well. For purposes of the SP method, we had wanted to find groups of households that were far enough apart so that households in comparison group clusters would not be easily aware of subsidies being received by households in the treatment group clusters. However, a quick review in Google Maps of the areas we will be studying suggested that there may not be many clusters of around 25 buildings that are clearly separated from other groups of buildings. Rather, our Google Maps search suggested that there are many groups of buildings that are much larger and a few that are much smaller.

For the SP method, we also wanted to find large numbers of clusters of households that were small enough so that we could reasonably afford to subsidize them all. We now suspect that there may be a sufficient number of small PSUs to satisfy this condition. In 1988, Tanzania had 7,812 PSUs (or Enumeration Areas) in their NMS of PSUs. Today, the population is around 40 million people. Assuming that the number of PSUs has not changed very much and that the number of people per household is around five, this implies an average of around 5,000 people and around 1,000 households per PSU. This includes urban PSUs, which are probably far larger on average than the peri-urban and rural ones that would be the focus of our study. A quick web search suggested that there are a large number of villages in Tanzania with fewer than 200 households (fewer than 1,000 people). The HBS 2001 data suggest that about two-thirds of the households in rural areas in the regions in our study have mud walls and thatched roofs. This means that in a village of 200 households only about 66 of them would be eligible for subsidies based on their building materials. Many of these households would not be within 30 meters of the electricity line (or proposed line) and thus would be excluded for that reason. Finally, since many PSUs are sub villages rather than villages they may be even smaller. Therefore, for our study we may be able to find large numbers of PSUs where only a small number of households in each PSU would be eligible for electricity subsidies.

While there may be a sufficient number of small PSUs in the rural areas of our study, there may not be a sufficient number being served by the line extensions funded by MCC which are often in peri-urban areas. This possibility means that we might not be able to afford to subsidize all eligible households in a reasonably large sample of PSUs for our treatment group. In that case, it may be possible to offset the size of the PSU by targeting the subsidies to a subset of the eligible households. Indeed, good targeting could reduce the number of households we would need to subsidize and survey. We could then focus the study more on those households most likely to be affected by the intervention. However, it is not clear at this point whether we will be able to develop an acceptable method of targeting to achieve this goal.

An advantage of using the PSUs is that we may be able to select them using the 2002 census data. This will work well if we can obtain sufficient 2002 census data by PSU and maps of those PSUs. If we cannot obtain the maps for the 2002 PSUs we may still be able to use them if the PSU boundaries have not changed significantly since that time (there were large changes in earlier years) and we can obtain good maps on the current PSUs. If there were large changes in the PSUs since 2002, it may still be possible to obtain relevant data from the National Bureau of Statistics on the new PSUs. The data we would need include the approximate number of households and their geographic locations. If we are not able to obtain complete data on any set of PSUs we may need to explore alternative methods of selecting our sample.

Our current plan is to obtain 2002 Census data and any more recent data on PSUs from the National Bureau of Statistics to select PSUs in the subproject and other comparison areas with relatively small numbers of households likely to be eligible. We would then use means testing to select appropriate households within those PSUs. This would allow us to both focus our study on the most appropriate households and limit the number of households per PSU. We could then afford to provide connection subsidies in a reasonably large number of PSUs.

Assigning groups of households, rather than individual households, to the treatment and comparison conditions reduces our statistical power. This reduction is greater when outcomes differ across these groups. More precisely, the reduction in power depends on the fraction of variance in the outcomes that is between groups, as opposed to within groups, of households. Consequently, in order to estimate statistical power, we needed to estimate the fraction of variance between PSUs. We did this using HBS 2001 data for the six regions where we will be conducting the D&E part of our study.

The variation between PSUs was between 9 and 13 percent of the total variation in outcomes we looked at (household income and assets). We use these estimates for our power calculations. If the actual variation between PSUs used in our analyses turns out to be greater, our statistical power may be somewhat less. If the variation between PSUs turns out to be smaller, our statistical power will be greater. In rural areas, these percentages are lower (five and eight percent). Thus, compared to looking only at rural areas, our power calculations are somewhat conservative.

All of our power calculations were done based on 80 percent power, five percent significance levels, two-tailed tests, and normally distributed outcome variables. We adjusted our power for the estimated variation between PSUs (also known as clustering), for the fact that the treatment and comparison groups were not always equal in size, for the possibility that covariates will explain some of the variation at the cluster level, for the fact that RD studies generally require about three times as much data as comparable random assignment studies, and for non-response at both the PSU and household levels.[4] We also assumed that there will be around $2,400,000 available for data collection covering two rounds (baseline and follow up), with a $80,000 fixed cost per round, $1,000 per cluster of households (or enterprises) per round and $110 per household (or enterprise) per round.

Power Calculations Results

Table B.1 presents our results. There are six sets of columns. The first set presents results for using the RD method. The second set shows results for the DID "early-late" method (DID-EL). The third set gives results relevant to the DID "no-treatment comparison group" (DID-CG) method or the SP method on its own, the fourth set (just one column) has estimates for the enterprise survey (which would also use one of the DID methods), the fifth set shows results for using the SP method in conjunction with one of the other methods, and the last set of columns presents results for the Cable IT study method.

The first four sets of columns are all relevant for estimating the impacts of line extensions. The treatment group is supposed to receive line extensions and the comparison group will not. If we implement the subsidy method and use one of these methods then some of the treatment group for estimating the effect of line extensions will also receive subsidies. In that case the power calculations below are relevant for estimating the combined effect of the line extensions and the subsidies that are available for some fraction of the treatment group.

RD Results: The first column in this set shows the ideal power we would have using the RD method if 200 subprojects were evenly distributed between the treatment and comparison conditions and if funding decisions were made based entirely on the revenue/capital expenditure ratio.[5] The third column shows our current understanding that only about 13 projects with low

---

[4] Schochet, Peter. "Technical Methods Report: Statistical Power for Regression Discontinuity Designs in Education Evaluations." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2008.

[5] We have seen lists with over 240 subprojects. However, we suspect that many of these will not be usable for our study for various reasons. For example, we may not be able to obtain complete information on the locations of the lines for all subprojects. For these calculations, we are assuming that 200 will be usable.

income/capex ratios will be left unfunded and we will have the same number of PSUs per subproject (on average) in both the treatment and comparison samples. Under the ideal condition (V1), the statistical power appears reasonable—we would be able to detect effect sizes as small as 0.29 even with conservative assumptions (13 percent of variance between PSUs and no covariates in our model). This would be like shifting income up from the median (50th percentile) to around the 61st percentile of the distribution. Alternatively, it would be like shifting the mean of a binary variable (such as having a radio) from 50 percent to around 64 percent.

The second column (V2) shows what would happen if we did our full follow-up survey in the subprojects that could be used for the RD method. This would mean dropping the 84 PSUs that were in non-project areas (see Table B.1) in the baseline survey and replacing them with 84 additional PSUs from the unfunded subproject areas. As shown in column V2 this reduces the minimum detectable effects by around 17 to 20 percent, a moderate gain in precision.

Unfortunately, the ideal situation illustrated in columns V1 and V2 does not appear likely at this time. Our current understanding is that there are only around 13 subprojects that are not funded out of 20 below the line that have low income/capital expenditure ratios. In addition, some of those above the line were funded. In net there is only a 58 percentage point difference in the rate at which subprojects above and below the line are funded. This means that we need to adjust our MDEs up by a factor of 1/0.58 to estimate the effect of actually having a subproject funded. The net effect is that under those conditions (V3), we might only be able to detect impacts with effect sizes of around 0.69 depending on the conditions. This would mean that income would have to shift up from the median to around the 75th percentile of the distribution, or the fraction of households with a radio would have to shift up from around 50 percent to around 85 percent. While such large impacts are possible, they may be somewhat unrealistic for many outcomes of interest. In addition, it should be noted that many of those 13 unfunded projects below the line may not be usable for our study if, for example, we cannot get good information on where those lines were supposed to be. This would further reduce our statistical power.

We understand that funding may be cut substantially which could result in an increase in the fraction of projects not funded. This could increase the statistical power. However, as discussed above that many of the projects below the current line were funded and many above the line were not funded. In other words there was a great deal of cross-over between the treatment and comparison groups. The net difference in the fractions funded is only around 0.58.[6] Columns V1 and V2 in Table B.2 are based on the assumption that all projects above the line are funded and none of the projects below the line are funded. This means that the estimated effects of being funded would have to be about 70 percent larger than those shown in columns V1 and V2 of Table B.1 in order to be detectable if the level of cross-over remains similar even after the

[6] Around 35 percent of the projects below the line were listed as being funded and around 93 percent of those above the line were listed as being funded, implying a difference of 58 percentage points.

TABLE B.2

Power Estimates for Tanzania Electricity Study Design Options

		Numbers of Units													
		Line Extensions										Subsidies with another study		Cable	
		RD			DID-EL			DID-CG or Subsidies Alone				w/RD or DID	w/DID-CG	IT	
		V1	V2	V3	V1	V2	V3	V1	V2	V3	Enterprise	V1	V2	V1	V2
Treatment	Households	3011	3200	5322	2048	903	2891	3011	3011	2164	32	1506	2891	na	na
Group	PSUs/Time Points	200	200	333	136	60	192	200	200	272	6	100	192	36	72
	HH/PSU	15.1	16	16	15.1	15.1	15.1	15.1	15.1	8.0	5.0	15.1	15.1	na	na
	Subprojects	100	100	180	68	30	30	na	na	na	na	na	na	na	na
	PSUs/Subproject	2.0	2.0	1.8	2.0	2.0	6.4	na	na	na	na	na	na	na	na
Control	Households	1506	2944	591	1988	903	2891	2770	1506	2164	32	1506	2891	na	na
Group	PSUs/Time Points	100	184	37	132	60	192	184	100	272	6	100	192	36	72
	HH/PSU	15.1	16	16	15.1	15.1	15.1	15.1	15.1	8.0	5.0	15.1	15.1	na	na
	Subprojects	100	100	20	66	30	30	na	na	na	na	na	na	na	na
	PSUs/Subproject	1.0	1.8	1.8	2.0	2.0	6.4	na	na	na	na	na	na	na	na
Level of Assignment/Clustering											PSU			Time	
Totals	Households	4517	6144	5913	4035	1807	5782	5782	4517	4327	65	3011	5782	na	na
	PSUs	300	384	370	268	120	384	384	300	544	13	200	384	72	144
Estimated Survey Budget		$1,753,750	$2,279,680	$2,200,000	$1,583,750	$797,500	$2,200,000	$2,200,000	$1,753,750	$2,200,000	$200,000	$1,222,500	$2,200,000	na	na
% Var Between	% explained by Covariate	Minimum Detectable Effects													
0%	0%	0.189	0.152	0.447	0.120	0.179	0.100	0.100	0.121	0.116	0.775	0.125	0.090	0.660	0.467
9%	0%	0.263	0.218	0.631	0.159	0.238	0.133	0.133	0.159	0.143	1.010	0.174	0.126		
	50%	0.231	0.189	0.545	0.139	0.211	0.117	0.117	0.140	0.134	0.960	0.153	0.109		
13%	0%	0.287	0.237	0.686	0.173	0.256	0.145	0.145	0.174	0.149	1.060	0.190	0.137		
	50%	0.244	0.199	0.575	0.148	0.223	0.123	0.123	0.150	0.137	0.980	0.162	0.115		

Notes:

Assume two-tailed tests, 80% power and 5 percent statistical significance.

For the DID methods assume 90% response rates for villages and 60% response rates by households within responding villages.

For the RD and SP methods assume 95% response rates for villages and 70% response rates by households within responding villages.

For the RD V3 option we assume a 57% difference in line extention rates between the treatment and comparison groups. For all other options we assume a 100% difference.

For the Enterprise survey we assume response rates of 0.90 at the village and enterprise levels.

% explained by covariate refers to covariates measured at the cluster level, such as the mean value of baseline responses to survey questions.

PSUs are defined using Tanzania Census enumeration areas, which are villages and subvillages in rural areas and census units in urban areas.

Adjusts for RD Design (assumes a design effect of 3) and unbalanced sample when appropriate.

Numbers do not add up exactly because of rounding.

RD=Regression Discontinuity	**Assumptions regarding survey costs**	
DID-EL=Difference in Difference Early/Late	$ per household	$110
DID-CG=Difference in Difference Comparison group	$ per cluster	$1,000
SP=Subsidy Pilot Study	$ fixed costs	$80,000
IT=Interuppted Time Series	Rounds of Survey	2

number of projects funded is reduced.[7] Thus, in order to make the RD method feasible we would need to see both a shift in the fraction of projects below the line towards 0.5 and a greater difference in the fraction of projects funded between those above and below the line than is true for the current list of projects we have.

DID-EL Results. The next set of three columns shows minimum detectable effects using the DID-EL method. The major issue here is figuring out how many sub-projects we will have to drop in order to leave enough time between when the early and late projects are implemented. We suspect that it will take at least one year between the two sets of projects for us to be able to observe reasonable impacts. This means that projects completed within the last year at the time of our follow up survey may not be useful for our study, as they could not be easily included in either the treatment or comparison condition. It is not clear how many projects we will lose from our study because of this constraint. In the table, we present estimates based on the assumptions that we will lose either 66 or 140 projects because of this limitation. If we lose only 66 projects, our statistical power could be noticeably better than the best possible power we would have for the RD method—in particular, we could be able to estimate impacts as small as a 0.17 effect size even under the more conservative assumptions (see column V1 in this set). Even if we lose 140 projects, the minimum detectable effect size would still be around 0.26 (see column V2), almost as good as the best we could do for the RD design.

If we did not use the flexible design option described above and we had good information on when the subprojects would be conducted, we could obtain a noticeable increase in statistical power as illustrated by column V3 in this set. The MDE would drop to only 0.145 under even the more conservative assumptions. Greater gains could be obtained if there were fewer subprojects in the middle group.

DID-CG or Subsidies Alone Results. The third set of columns (under DID-CG or Subsidies Alone) shows our statistical power if we use the DID-CG method or if we use the SP method alone (that is, without studying the effects of line extensions). Under these conditions, our statistical power improves because we are no longer constrained to using subprojects as our unit of assigning households to the treatment and comparison conditions. Rather, we can assign smaller units, such as PSUs. If we use 384 PSUs (all of the PSUs listed in the TOR) we can estimate impacts of around 0.145 in effect size under the conservative assumptions (see V1). If we drop the non-project PSUs and therefore drop the number of PSUs to 300, the MDE rises moderately, to 0.174. If we reduce the number of households per PSU and increase the number of PSUs to 544 the MDE moves up slightly to 0.149. Thus, comparing columns V1 and V3 it appears that reducing the number of households per PSU further may not provide additional power.

[7] Schochet (2008) shows that one needs to adjust the minimum detectable effect by the factor 1/0.58 when the difference in participation rates between the treatment and comparisons groups is 58 percentage points compared to when there is a 100 percent point difference. This gives an adjustment of 1.7 (=1/0.58) meaning the estimated effect must be 70 percent larger to be detectable.

Our calculations for both the DID-EL and DID-CG methods assume that we will be able to match the treatment and comparison groups relatively well based on baseline characteristics. If this is not the case, the minimum detectable effects will likely be larger than reported here because the background variables used in our analyses will be correlated with the treatment variable and thus reduce the amount of independent variation in that variable.

Enterprise Survey: The enterprise survey will be analyzed using one of the DID methods because there will not be sufficient sample size to estimate impacts with any precision using the RD method. Thus, the power calculations are based on a similar set of assumptions. We present only one set of power calculations based on a very small sample size (65 businesses) given that we were asked to keep the budget for this survey very small. Unfortunately, the cost data we collected suggested an $80,000 fixed cost per survey instrument. This means that if we conduct both a baseline and follow-up enterprise survey it might cost $160,000 just to put the survey in the field, without collecting any data. The result is that the resulting sample size is very small and the statistical power is fairly low. In particular we estimate that given this budget constraint the estimated impacts would have to be close to 0.78 even under fairly weak assumptions (for example no clustering). It we determine a cheaper way to conduct the enterprise survey, it may be possible to obtain larger sample sizes and increase the statistical power of this part of our study.

Subsidies with Another Method Results. The fifth set of columns shows calculations for using the SP method in conjunction with one of the other methods. We assume that this is likely since it would result in major savings in data collection costs compared to conducting the studies separately. The first column in this set shows what the power would be if we were to do the SP method with the V1 versions of the RD or DID-CG methods, each of which has 200 PSUs in the treatment group with electrical lines. For the SP method, these groups would be split evenly between the treatment and comparison group sets. The resulting power would enable us to detect impacts with effect sizes of around 0.190. The second column shows what would happen if we were to use all 384 PSUs covered in the TOR for the SP method. The MDE drops to 0.137 even under the most conservative assumptions.

There is a potential tension between focusing the baseline survey on the subprojects most likely to be funded, which could reduce risk for the SP method, and using the flexible design option which reduces risk for estimating the impacts of line extensions but involves doing the baseline survey in all (or almost all) subprojects. If all (or almost all) of the subprojects that are supposed to be funded are funded then there will be no tension between these two options. If only a subset are funded then we may end up needing to conduct the follow-up surveys in a different set of PSUs than the baseline survey. This could reduce statistical power when estimating both the impact of line extensions and the impacts of subsidies. For example, suppose that we were planning to have 100 subprojects funded and conduct the baseline survey in two PSUs in each of those subprojects. However, shortly after conducting the baseline survey we discover that only 30 of those subprojects will be funded. We would then have a sample size of only 60 PSUs (two per subproject for 30 subprojects) for the SP method. We could augment this by adding additional PSUs in these subprojects to get back up to the original 200 PSUs. This would still result in somewhat less statistical power than originally planned for the SP method as we would not have baseline survey data for all of these PSUs. There would also be some loss in statistical power for estimating the effect of line extensions. If we were using the RD method then the resulting sample would probably not be optimal because the number of PSUs in the funded projects would be much higher than in the unfunded projects. This probably wouldn't have a major impact on power. The effects on power would be larger if we were to use one of the DID methods because we would not be able to use the data on these new PSUs as we would have no baseline data on them.

There are two types of risk for the SP method. First, the number of subprojects funded may turn out to be less than expected. Second, the selection of which subprojects are funded may change. The flexible design option might increase risk for the SP method if the number of subprojects drops. However, the flexible design option would actually reduce risk if the selection of which subprojects will be funded changes. Suppose, for example, that we focus the baseline survey on only the 30 subprojects most likely to be funded and later discover that only 15 of them are going to be funded and that another 15 that were not as high on the list are being funded instead. In that situation the flexible design option might actually be preferable to the strategy of focusing on the subprojects most likely to be funded.

Regardless of what changes occur one needs to keep in mind that in order for a PSU to be useful for the SP method we need to have sufficient time to decide whether or not to put it into the treatment or comparison group and, if that PSU is in the treatment group, we need to be able to have time to offer the subsidies. Thus, while the flexible design option does give some additional time for making these decisions, the amount of additional time is limited as we will likely need to be able to make the subsidy offers at least one year before the follow-up survey is conducted.

Effects of Connections Versus Subsidies. The MDEs for estimating the impacts of connections will be at least as large and probably much larger than the MDEs for estimating the effects of subsidies. To see why note that if all households that were offered subsidies connected and none of the households without subsidies connected then the effects of subsidies and connections would be the same. In reality, this is not likely. Some households with subsidies will not connect and many households without subsidies will connect. Consequently, the impact of subsidies is likely to be smaller than that of being connected. For example, if 55 percent of households with subsidies connect and 5 percent of households without subsidies connect then we would expect the effect of subsidies to be about half as large as the effect of being connected (55 percent minus 5 percent). This would mean that in order for us to detect an impact of being connected using our methods, the estimated impacts would need to be about twice as large as those found in Table B.1. More generally, the MDEs for the subsidy method in Table B.1 are for estimating the impacts of being offered a subsidy. To estimate MDEs for estimating the impacts of being connected we will need to multiply these by the ratio of one over the difference in connection rates between the treatment and comparison groups.

In the example above we assumed that only 5 percent of the households not offered subsidies connect. These are the "always-connect" households. As discussed earlier these households benefit from the subsidy because of its monetary value and not because they are getting electricity. Thus, in order to estimate the effects of being connected to electricity these households must constitute at most a small fraction of our total sample.

As noted above, in order to estimate the impacts of being connected we would first need to estimate the difference in connection rates between the treatment and comparison villages. When doing this we would need to consider both the household level decisions to make use of the subsidies and the possibility that not all villages in the treatment group will be offered the subsidies. Villages may not be offered subsidies if, for example, the individuals that were supposed to tell the village about the subsidy offers were unable to do so for logistical and/or financial reasons. They might have been unable to find the village, tried to visit the village when the road was washed out by rain, or run out of funding for their part of the project before contacting the village.

The estimates of being in the treatment group compared to the comparison group (or being offered a subsidy) are sometimes referred to as Intent to Treat estimates. This highlights the fact that these estimates identify the impact of being in the treatment group (being in a PSU that was supposed to be offered a subsidy) but not necessarily that of being treated (actually getting connected). In contrast, the estimates of getting an electricity connection are often referred to as Complier Average Causal effects to highlight the fact that they are only relevant for the part of the sample that gets connected to electricity if and only if they are in the treatment group (that is the "Subsidy-Connect" households).[8] Non-compliers are those households that would have been connected to electricity regardless of whether or not they were in the treatment group (the "Always-Connect" households) and those households that would not connect regardless of their treatment status (the "Never-Connect" households). The behavior of non-complying households is unaffected by treatment, so comparisons of a treatment and comparison group cannot provide information on the effects of electricity for them.

Cable IT Results. The last set of two columns presents estimates of the amount of power we would have for the Cable IT method as a function of the number of time periods of data we have to study. If we have three years of monthly data from before and after the Cable project is completed, this would give us the ability to detect an effect size as small as 0.66. This means that electricity reliability or quantity would have to increase such that in a typical month it was at least as good as it was in the top quarter of months from before the Cable project was implemented. While this would be a large shift for many of our other outcomes, it may not be an unreasonable expectation for the effect of the Cable project directly on electricity reliability and quality. If we could double the number of time periods of data, either by obtaining data for shorter periods of time (say two-week or one-week periods) or a longer time frame (say six years instead of three), our statistical power would increase as shown in the last column of Table B.2.

Sample Attrition. The power calculations above adjust for attrition at both the PSU and household level. Based on previous experience we assume that 95 percent of PSUs selected will participate in our study; within those PSUs we expect a response from 70 percent of households.

[8] Schochet, Peter Z. and Hanley Chiang (2009). Estimation and Identification of the Complier Average Causal Effect Parameter in Education RCTs (NCEE 2009-4040). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

These are the response rates that matter for the RD and SP methods as we do not need baseline data for those methods. For the DID methods we need both baseline and follow-up data. Consequently, for those studies we assume lower response rates—90 percent at the village level and 60 percent at the household level. Given the smaller sample sizes involved in the enterprise survey, we assume greater effort will be exerted to achieve high response rates there so we assume 90 percent response rates at both the village and enterprise levels for this part of the data collection effort (including both baseline and follow-up surveys).

Implications for Subsidy Levels. These power calculations suggest sample sizes of PSUs and households per PSU. In Table B.3, we explore the implications of these assumptions for the possible subsidy levels that could be offered. The first three set of columns shows what happens as we increase the number of PSUs from 50 to 150. The subsidy rate would fall from around 60 percent to around 20 percent.

The first set of columns was based on an assumption of a 100 percent take-up rate. The take-up rate may be somewhat lower. However, if we assume a lower rate then we run a risk of going over budget. The second set of columns shows what happens if we assume a 40 percent take up rate but the actual take-up rate is higher. If the take-up rate rises to 80 percent, the costs of the subsidies double. This may be acceptable if we implement the subsidies close to the end of a fiscal year. As we understand it we could use $300,000 per year for these subsidies. If we discover that more households connect than expected then we could ask for additional funding in the later fiscal year to cover the short-fall. However, even this method would have limits. As shown in the third column of the second set, if the take-up rate more than doubles, the costs will more than double and put us over budget. One way to avoid going over budget while still starting with an assumed take-up rate below 100 percent is to assume a take-up rate of at least 50 percent and plan to use only half the budget.

In the third set of columns we show what could happen if the number of households per villages changes from what we assume. Again, clearly there is a budget risk. If we can develop a firm method of identifying which households are eligible for subsidies we may be able to avoid this risk. The last set of columns shows that our risk would be much greater if we both assume a take-up rate that is too low and a number of households per cluster that is too low.

Conclusions Regarding Power. While we expect to have sufficient power using the DID-CG method, each of the other methods has significant risks. The IT method (for evaluating impacts of the Cable project on electricity reliability and quality) and analyses based on the Enterprise Survey will likely have the largest MDEs, though these may be sufficient for the IT method given the likely large impacts on those outcomes. The DID-EL method is at risk if we lose too many projects due to the need to select only projects that end early and start late. The RD method is at risk as it appears that there are not enough subprojects that will be unfunded. Also, even if there are more unfunded projects it appears that there is a great deal of cross-over between the treatment and comparison groups which could further weaken our ability to estimate impacts using that method. Finally, even the SP method is at risk, as there may not be enough small PSUs in the subproject areas to enable us to obtain a good sample size and still be able to afford to subsidize all eligible households.

TABLE B.3

EXPLORING OPTIONS FOR USE OF SUBSIDY BUDGET FOR TANZANIA ELECTRICITY PROJECT EVALUATION

	Vary PSUs			Vary Take-up Rate			Vary HH/PSU			Vary HH/PSU and Take-up Rate	
PSUs	50	100	150	100	100	100	100	100	100	100	100
HH/PSU	25	25	25	25	25	25	25	50	75	25	50
Take-up Rate	100%	100%	100%	40%	80%	100%	100%	100%	100%	50%	100%
Total Treatment HH	1250	2500	3750	1000	2000	2500	2500	5000	7500	1250	5000
$ per HH	$240	$120	$80	$300	$300	$300	$120	$120	$120	$240	$240
Budget	$300,000	$300,000	$300,000	$300,000	$600,000	$750,000	$300,000	$600,000	$900,000	$300,000	$1,200,000
% subsidy	60%	30%	20%	75%	75%	75%	30%	30%	30%	60%	60%

D. CRITICAL ASSUMPTIONS

To draw conclusions based on our results, we will need to make certain critical assumptions. Given the limitations of many of our study designs, we plan to present many of our results cautiously, rather than rely heavily on these assumptions. For example, for the ITS study we do not expect to draw strong conclusions regarding the impact of the Cable project, since other factors that affect electricity reliability and quality may have changed at the same time that the Cable project was implemented. A critical assumption needed to draw conclusions based on the ITS design is that no other large changes occurred during this time that would impact the outcomes we are measuring. A similar assumption holds regarding the PPC study, and we plan to be particularly cautious when describing results from that study.

In contrast to our results based on the IT and PPC methods, we hope to be able to draw somewhat stronger conclusions about the impacts of the MCC-funded electricity projects based on results from the RD, DID, and SP methods.

The results from the RD method can be interpreted in two ways depending on how the results turn out. If we find that the line extensions appear to have no impact on roads or water then we can use the RD method to estimate the direct effect of having improved access to electricity directly on household outcomes. On the other hand, if we find that there are significant changes in water and road access associated with line installations, perhaps driven by government decisions to improve water, roads, and electricity at the same time, then any changes in the household outcomes may be due in part to all of these changes and not just to the direct effects of the new electric lines.

Another assumption needed to justify the RD method is that we have correctly parameterized the relationship between the outcomes and the score variable (in our case the revenue/capital expenditure ratio) in the areas just above and below the cut-point. If we use this method we will test our estimates to see how robust they are to variations in how we parameterize these relationships.

The DID methods require a similar but somewhat stronger assumption than the RD method. In particular, we need to assume that in the absence of the intervention, the treatment and comparison groups would experience similar changes in their outcomes. This would be violated, again, if the treatment group received more new interventions of other types (such as roads and water) than the comparison group households.

The strongest method we have is the SP method when estimating the impacts of subsidies for households with line extensions. For that method and question, we only need to assume that nonresponse to the follow-up survey is not biasing our estimates—in other words, we need to assume that the nonresponding households in our treatment group and comparison group had similar differences in outcomes as those that did respond. Similar assumptions must also be made for each of the other study methods.

We are also hoping to use the SP method to estimate the impact of having subsidies and line extensions compared to no line extensions. In that case the SP method is similar to the DID methods in terms of the required assumptions since the comparison group will have to be selected using some sort of matching method (DID-EL or DID-CG).

The final use of the SP method is to estimate the impacts of being connected. In order to justify this we need to assume that the treatment/comparison group differences are unaffected by differences in outcomes for the "Always-Connected" households. If these households are a small fraction of the households in our study or if the subsidy is a small fraction of household income for these households then this assumption may be well justified. We will be able to test the latter assumption by calculating the average income of the households in the comparison group that connect and comparing that to the size of the subsidies we provide.

Our study methods are also based on the assumption that we will be able to draw a good sample of clusters of households. One option is to rely on the data provided by the survey firm as they walk the line. This could entail a great deal of walking on their part. To reduce the amount of walking we are proposing to use data by PSU, including the numbers of households in each PSU and, if possible, more detailed information on the numbers of households that are not thatched or with mud walls. These data could come from the 2002 Census or more recent information that the NBS may have. Other household characteristics (such as the fraction with radios) would also enhance this method. Then, we would use the current PSU maps to locate these PSUs. This is based on the assumption that both of these pieces of information are available and that the current PSUs are reasonably similar to the old ones. If either of these assumptions is not true, we would need to rely on the survey firm to collect data and provide them to use when they walk the line. If the survey firm does this work we will likely end up with somewhat reduced statistical power as it is likely they will not collect data for as many potential PSUs as we could get from the census data and therefore not be able to do as good of a job at matching the treatment and comparison group PSUs. In terms of our statistical power this can be thought of as reducing the amount of variation explained by the covariate.

The new flexible design option also has limitations. In particular it still requires that we have a good idea where the lines will be before conducting the baseline survey. ESBI may provide sufficient information to make this determination when they write their tender for local firms that are installing the new lines, perhaps in August 2009. If that is not the case, we may need to wait until the local firms have produced their work plans. This could significantly delay the baseline and follow-up surveys. This could be a problem for any of our current research methods.

E. NEXT STEPS

Following is a list of next steps and a proposed schedule for this project.

- Cable Project – Design
  - Obtain additional baseline data from ZECO (or TANESCO) on the outcomes of interest for a large number of months (preferably at least 36). Those data would enable us to conduct a power analysis that could be used to determine how many data points would be needed in order to estimate impacts of a

certain size. Currently we have only about one-month of data from TANESCO and no data directly from ZECO.

- Review materials from the Johns Hopkins graduate students conducting a power sector research project in Zanzibar and explore possible collaboration.

• D&E project – Design

  - Update power calculations when additional data become available.
  - Determine when and how to choose which method will be used to estimate the impacts of line extensions (RD, DID-EL, or DID-CG). This could probably be done around the time of the baseline survey if we use the "flexible option" discussed above.
  - Determine when and how to decide if we can use the SP method.
  - Obtain ESBI's contract and work plans so that we understand exactly what information (for example, maps, GPS coordinates, and so on) they are likely to provide and if possible obtain samples of these data so that we can assess their likely usefulness.
  - Obtain updated list and timelines for the funded subprojects.
  - Decide on an appropriate approach for identifying comparison households.
    - RD – Determine method to estimate where the line would have been extended in comparison group areas.
    - DID-EL – Determine whether we will know in advance where the line will be placed in the areas that will have line extensions later and thus will be in our comparison group. If not, determine method to estimate where the line will be extended.
    - DID-CG – While we will use propensity matching to identify the PSUs, we need to determine how we will locate households within these PSUs. One idea is to identify the locations of major roads and use those as a proxy for identifying the likely locations of line extensions.

• Data Collection

  - Contact the Department of Land Registry, Government of Tanzania, about the price and availability of digital maps (roads, rivers, hospitals) that could be used to inform development of a sampling frame.
  - Follow up with NBS on availability of data on PSUs including number of households and digitized maps.
  - Determine how NBS updates information regarding enumeration areas for their 2012 census.

- One of the outcomes highlighted in Dinkelman (2008) is the amount of time women spend collecting firewood.[9] It may also be useful for our studies to have other information on how people use their time. For this reason we will likely want to include time-use questions in our household surveys. We need to investigate how this can best be accomplished.
- Obtain data from TANESCO on how many people pay tariffs per month for new lines in the months immediately after the lines are installed so that we can determine how long we need to wait after lines are installed before we are likely to see changes in outcomes.

- Refine plans for implementing subsidies

    - Determine how much the subsidy should be, based on budget and design considerations, document reviews, and discussions with other donors.
    - Learn more about the Torero study. How did Torero implement the subsidies? Did he use a local consultant to oversee implementation? What challenges did he face?
    - Learn more about the Swedish International Development Cooperation Agency (SIDA) studies. What implementation challenges did they face?
    - Obtain the REA cost of services study to be released in July 2009.
    - Get information about the new proposed tariffs in August 2009.
    - Consider the feasibility of clusters smaller than a village.
    - Consider whether it is possible to avoid subsidizing households that are not being surveyed .

[9] Dinkelman, Tayrn. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." Working paper. Princeton, NJ: Princeton University, 2008.

TABLE B.4

PROPOSED TIMETABLE FOR PHASE 2 TANZANIA ENERGY SECTOR EVALUATION STUDIES

Actual Date	**Activity/Deliverable**	**Date in Proposal**
	OPTION YEAR 1	
July 24, 2009	Beginning of Option Year 1	
August 9, 2009	Responses due for TOR for data collection (6 weeks)	
August 2009	ESBI issues tender (RFP) for implementation of T&D interventions	
August 10-14, 2009	Review of applications by Cheda, Anne, and other MCC/MCA-T staff members. MPR staff originally slated to participate, but MPR cannot be a voting member. (At present, no trip is budgeted for MPR for this activity)	October 2008
August 17, 2009	Selection of contractor for local data collection and notification	
August 24, 2009	End of Phase 1 activities	
	Start of Phase 2 activities	
September 14, 2009	Billing rates increase	
September 30, 2009	MCA-T signs contract(s) with local data collection firm	
October 12, 2009	Minki travels to Tanzania to start work with local research firm on sample selection plan. There may be some follow-up tasks related to this, depending on the capacity of the firm.	January—May 2009
October 28, 2009	Sample Selection Plan received from local data collection firm	
November 4, 2009	MPR provides comments on Sample Selection Plan	
November 11, 2009	Draft English survey instruments and Pretest Plan received from local data collection firm	
November 18, 2009	MPR provide comments on draft English survey instruments and pretest plan	
November 25, 2009	Back-translations of Kiswahili and Kiswahili survey instruments received from local data collection firm.	
December 2, 2009	MPR provides feedback on back-translations and suggests revisions to instrument	
December 14-25, 2009	Pretest occurs in Tanzania – Kathy travel to participate	

TABLE B.4 (*continued*)

Actual Date	Activity/Deliverable	Date in Proposal
December 23, 2009	Survey Implementation Plan received from local data collection firm.	
January 15, 2010	Pretest Report and Pretest versions of instruments (English and Kiswahili) received	
January 29, 2010	MPR provides comments on Pretest Report, suggested changes to survey instruments, Survey Implementation Plan	
February 12, 2010	Revised English, Kiswahili and back-translated questionnaires received from local data collection firm.	
February 19, 2010	MPR provides comments on instruments	
February 26, 2010	Final versions of instruments (English, Kiswahili) and Survey Implementation Plan received from local data collection firm.	
January 2010	Application deadline for proposals to be submitted to ESBI for implementation of T&D interventions	
February/March 2010	Interviewer Training is held in Tanzania (Minki travels to participate/Kathy arrives at end for data entry training and to oversee the beginning of baseline data collection)	
March 2010	Award of contract(s) to local firm(s) that will implement the T&D interventions	
March 2010	Baseline data collection begins (This assumes that the tender issued by ESBI in August clearly states where the power lines will be placed. If this information is not available, data collection will have to wait for those maps.)	May 2009
June 2010	Contract mobilized to local firm(s) that will implement the T&D interventions	
July 23, 2010	Option Year 1 ends	
OPTION YEAR 2		
July 2010	Local firm that is implementing the T&D interventions has developed their first version of their workplans and has information (GPS coordinates and maps) about which subprojects will be implemented. The firm has also developed their schedule.	
July 2010	MPR uses information from the local implementing firm to make final design decisions and produces memo to update Final Design Report.	
August 2010	Baseline data collection ends	
September 2010	Baseline data entry ends	

TABLE B.4 (*continued*)

Actual Date	Activity/Deliverable	Date in Proposal
September 13, 2010	Impact Evaluation Master Contract expires	
September 14, 2010	Billing rates increase	
October 2010	Baseline data cleaning and creation of codebook by local data collection firm	
November 2010	Baseline data ready for analysis	
December 2010 January 2011	MPR analyzes baseline data and assigns subsidy and non-subsidy villages	
January 2011	T&D intervention implementation begins	
February 2011	Subsidy implementation begins	
March 2011	Duncan (or Minki) travels to Tanzania to oversee implementation of selection criteria for the beneficiaries of subsidies, if applicable	
April 2011	Data collection firm conducts subsidy verification visits	
July 23, 2011	Option Year 2 ends	
OPTION YEAR 3		
September 14, 2011	Billing rates increase	
September 2011	Denzel and Minki work with local consultant to put together plan to develop sampling frame and select hotels for PPC study in Zanzibar	
October 2011	Minki travels to Tanzania to work with local consultant to finalize plan for sampling frame, selection of hotels, how to approach hotels for PPC study in Zanzibar	
October 2011	Local consultant designs sampling frame, selects hotels, approaches hotels to participate in the PPC study	
October 2011	Denzel designs draft costing tool in Excel and MPR does quality assurance of costing tool for PPC study	
November 2011	Denzel travels to Tanzania to work with local consultant to start collecting costing data from hotels for PPC study	
December 2011 January 2012	Local consultant finishes collection of costing data for PPC study	

TABLE B.4 (*continued*)

Actual Date	Activity/Deliverable	Date in Proposal
January 2012	Zanzibar cable completed	
February-March 2012	Denzel analyzes pretest cost data for PPC study	
April 2012	Denzel produces memo on findings from baseline costing data collection for PPC study	
May 2012	T&D intervention implementation ends	
July 23, 2012	Option Year 3 ends	
OPTION YEAR 4		
September 14, 2012	Billing rates increase	
October 2012	Minki travels to Tanzania to plan sample selection and follow-up surveys plans with local data collection firm	
October 28, 2012	Follow-Up Sample Selection Plan received from local data collection firm	
November 4, 2012	MPR provides comments on Follow-Up Sample Selection Plan	
November 11, 2012	Follow-Up draft English survey instruments received from local data collection firm	
November 18, 2012	MPR provide comments on Follow-Up draft English survey instruments and pretest plan	
November 25, 2012	Back-translations of Kiswahili and Kiswahili survey instruments received from local data collection firm.	
December 2, 2012	MPR provides feedback on back-translations and suggests revisions to instrument	
December 14-25, 2012	Pretest occurs in Tanzania—Kathy travel to participate	
December 23, 2012	Follow-Up Survey Implementation Plan received from local data collection firm.	
January 15, 2013	Follow-Up Pretest Report and Pretest versions of instruments (English and Kiswahili) received	
January 29, 2013	MPR provides comments on Pretest Report, suggested changes to survey instruments, Survey Implementation Plan	
February 12, 2013	Revised English, Kiswahili and back-translated questionnaires received from local data collection firm.	
February 19, 2013	MPR provides comments on instruments	

TABLE B.4 (*continued*)

Actual Date	Activity/Deliverable	Date in Proposal
February 26, 2013	Final versions of Follow-Up instruments (English, Kiswahili) and Survey Implementation Plan received from local data collection firm.	
February/Mar 2013	Follow-Up Interviewer Training is held in Tanzania (Minki/Kathy travel to participate)	
March 2013	Follow-Up data collection begins (this assumes that the tender issued by ESBI in August clearly states where the power lines will be placed. If this information is not available, data collection will have to wait for those maps.)	
July 23, 2013	Option Year 4 ends	
OPTION YEAR 5		
August 2013	Follow-Up data collection ends	
September 2013	Follow-Up data entry ends	
September 14, 2013	Billing rates increase	
October 2013	Follow-Up data cleaning and creation of codebook by local data collection firm	
December 2013	Follow-Up data ready for analysis	
October 2013	Local consultant updates sampling frame for PPC study for post-test	
November 2013	Denzel travels to Tanzania to work with local consultant to start collecting costing data for post-test from hotels for PPC study	
December 2011 January 2013	Local consultant finishes collection of costing data for post-test for PPC study	
February-April 2014	Denzel analyzes pretest and post-test cost data for PPC study	
May 2014	Denzel produces PPC study report together with MPR	
January-March 2014	Data analysis for impact evaluation	
April 2014	Report writing and Final Presentation	
July 2014	Final Impact Evaluation Report completed	